Review Artikel: Psychometric Characteristics of the Culture Fair Intelligence Test Scale 2

Latar Belakang Penelitian

Pengukuran kecerdasan merupakan aspek penting dalam bidang psikologi dan pendidikan, digunakan untuk berbagai tujuan seperti diagnosis anak-anak, rujukan ke program pendidikan khusus, dan penanganan disabilitas tertentu (Tellegen & Laros, 1993). Salah satu tantangan utama dalam pengukuran kecerdasan adalah kemungkinan adanya bias budaya dalam tes yang digunakan. Nenty & Dinero (1981) menekankan bahwa pengalaman budaya individu dapat mempengaruhi respons terhadap butir tes, yang dapat menyebabkan bias dalam hasil pengukuran.

Untuk mengatasi masalah ini, Raymond B. Cattell mengembangkan Culture Fair Intelligence Test (CFIT) yang bertujuan untuk mengukur kecerdasan cair (fluid intelligence) tanpa dipengaruhi oleh latar belakang budaya atau lingkungan individu (Gregory, 2013). CFIT menjadi instrumen yang banyak digunakan dalam berbagai penelitian, terutama CFIT-skala-2 yang diperuntukkan bagi individu dengan kecerdasan rata-rata, termasuk anak-anak berusia 8 hingga 13 tahun.

Meskipun CFIT-skala-2 telah lama digunakan, alat ini belum mengalami revisi sejak tahun 1961. Oleh karena itu, ada kebutuhan untuk meninjau kembali karakteristik psikometriknya untuk memastikan bahwa alat ini masih relevan dan bebas dari bias yang dapat mempengaruhi hasil pengukuran. Studi ini bertujuan untuk mengevaluasi kembali karakteristik psikometrik CFIT-skala-2 dengan menganalisis tingkat kesulitan, daya beda, efektivitas distraktor, validitas konstruk, reliabilitas, serta menguji keberadaan Differential Item Functioning (DIF) yang dapat menunjukkan adanya bias dalam tes.

Tujuan Penelitian

Penelitian ini bertujuan untuk menganalisis karakteristik psikometrik dari CFIT-skala-2. Beberapa aspek utama yang diteliti meliputi:

- Tingkat Kesulitan dan Daya Beda: Mengevaluasi apakah tingkat kesulitan butir tersusun secara sistematis dari yang mudah ke yang sulit dan apakah daya beda butir sesuai standar.
- Efektivitas Distraktor: Mengukur seberapa baik pilihan jawaban yang salah dalam tes dapat membedakan individu dengan dan tanpa kemampuan yang sesuai.
- Analisis Diferensial Item Functioning (DIF): Menentukan apakah butir dalam CFIT-skala-2 memberikan hasil yang adil bagi semua kelompok berdasarkan jenis kelamin.
- Validitas Konstruk: Menggunakan analisis faktor konfirmatori (CFA) untuk menentukan apakah model pengukuran sesuai dengan konstruk teoretis.
- Reliabilitas: Mengukur konsistensi internal tes menggunakan koefisien Cronbach’s alpha dan reliabilitas konstruk.

Metode Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode psikometri. Responden terdiri dari 507 siswa (245 perempuan dan 262 laki-laki) dengan rentang usia 8 hingga 13 tahun (M=10,88; SD=1,35). Teknik pengambilan sampel dilakukan secara acak dari berbagai sekolah dan kelompok bimbingan belajar. Sebelum pengambilan data, peneliti memberikan informasi kepada guru dan wali siswa mengenai tujuan dan prosedur penelitian, serta mendapatkan izin dari orang tua.

Instrumen yang digunakan dalam penelitian ini adalah CFIT-skala-2, yang terdiri dari empat subtes memuat seri, klasifikasi, matriks, dan tipologi. Pengujian dilakukan dalam sesi kelompok, dengan penyampaian instruksi yang terstruktur dan batasan waktu untuk setiap subtes. Data dianalisis menggunakan teori tes klasik (CTT) untuk mengevaluasi tingkat kesulitan, daya beda, dan efektivitas distraktor. Selain itu, analisis teori

respons butir (IRT) model 2 Parameter Logistic (2PL) digunakan untuk menilai karakteristik butir secara lebih mendalam. Uji validitas konstruk dilakukan dengan CFA, sementara reliabilitas diuji menggunakan Cronbach's alpha dan reliabilitas konstruk.

Hasil Utama Penelitian

Analisis CTT menunjukkan bahwa tingkat kesulitan butir dalam CFIT-skala-2 bervariasi dari mudah hingga sulit, tetapi tidak tersusun secara sistematis dari yang paling mudah hingga paling sulit. Sebanyak 28 butir tidak memenuhi kategori sangat baik dalam daya beda ($p > 0,40$), yang berarti butir-butir ini kurang efektif dalam membedakan individu dengan kemampuan tinggi dan rendah. Sebanyak 47 (25%) dari 184 distraktor dalam tes ini dinyatakan tidak efektif, yang berarti banyak pilihan jawaban yang salah tidak berfungsi dengan baik dalam menggiring individu dengan kemampuan rendah untuk memilihnya.

Tidak ditemukan adanya DIF dalam CFIT-skala-2 ($Adj.p > 0,05$), yang menunjukkan bahwa tes ini tidak menunjukkan bias berbasis gender dan dapat digunakan secara adil bagi semua individu dengan tingkat kemampuan yang sama. Lebih lanjut, hasil CFA menunjukkan bahwa CFIT-skala-2 memiliki model yang sesuai dengan konstruk teoritisnya ($p < 0,001$; $RMSEA < 0,08$; $SRMR < 0,08$). Hal ini mendukung kesesuaian antara model teoritis dan data empiris. Berkaitan dengan reliabilitas, koefisien Cronbach's alpha sebesar 0,88 ($\alpha > 0,70$), menunjukkan bahwa CFIT-skala-2 memiliki reliabilitas internal yang baik. Namun, reliabilitas konstruk masih di bawah nilai yang dapat diterima ($CR < 0,69$), yang berarti bahwa beberapa aspek pengukuran dalam CFIT-skala-2 masih perlu diperbaiki.

Kesimpulan Penelitian

Berdasarkan hasil penelitian, CFIT-skala-2 masih memiliki beberapa kekurangan yang perlu diperbaiki. Susunan tingkat kesulitan butir dalam tes ini tidak terstruktur secara sistematis, sehingga bisa mempengaruhi

interpretasi hasil tes. Selain itu, daya beda beberapa butir kurang optimal, serta efektivitas distraktor tidak sepenuhnya memadai. Meskipun tes ini tidak mengandung bias berbasis gender dan memiliki validitas konstruk yang baik, reliabilitas konstruk masih perlu ditingkatkan.

Implikasi dari penelitian ini adalah perlunya revisi dan evaluasi ulang terhadap butir-butir dalam CFIT-skala-2 untuk meningkatkan akurasi pengukuran kecerdasan cair individu. Para psikolog dan biro asesmen psikologi disarankan untuk menggunakan CFIT-skala-2 dengan hati-hati, serta mempertimbangkan penggunaan instrumen lain sebagai pelengkap dalam asesmen psikologi. Secara keseluruhan, penelitian ini memberikan wawasan yang penting tentang karakteristik psikometrik CFIT-skala-2 dan menegaskan perlunya perbaikan dalam penyusunan butir agar lebih sesuai dengan prinsip pengukuran psikometri modern.

Keandalan dan validitas tes

Menurut penelitian ini, alat ukur Culture Fair Intelligence Test Skala 2 (CFIT-skala-2) memiliki validitas dan reliabilitas (keandalan) yang baik.

- **Reliabilitas**: CFIT-skala-2 memiliki koefisien reliabilitas Cronbach's alpha sebesar 0,88 ($\alpha > 0,70$), yang menunjukkan bahwa tes ini memiliki reliabilitas yang baik. Namun, reliabilitas konstruk berada di bawah nilai yang dapat diterima ($CR < 0,69$), sehingga perlu evaluasi lebih lanjut (Zahra, Widyastuti, & Ridfah, 2022).
- **Validitas**: Hasil uji validitas konstruk menggunakan Confirmatory Factor Analysis (CFA) menunjukkan model yang baik ($p < 0,001$; $RMSEA < 0,08$; $SRMR < 0,08$), yang mendukung kesesuaian antara model teoritis dan empiris. Hal ini menunjukkan bahwa tes ini cukup valid dalam mengukur konstruksi yang dimaksud.

Penelitian serupa oleh Fadillah (2021) juga menyoroti pentingnya analisis mendalam terhadap properti psikometri CFIT, khususnya terkait peluang guessing yang dapat mempengaruhi validitas tes

Kelebihan dan Kekurangan Penelitian

Secara aspek kelebihan, penelitian ini menggunakan metode yang komprehensif dengan pendekatan teori tes klasik (CTT) dan teori respons butir (IRT) untuk mengevaluasi karakteristik psikometri CFIT-skala-2. Kemudian, jumlah sampel yang digunakan cukup besar sejumlah 507 siswa yang membuat analisis psikometrik lebih akurat. Tes ini tidak menunjukkan adanya Differential Item Functioning (DIF), yang berarti tidak ada bias berdasarkan gender. Lebih lanjut, hasil penelitian ini dapat membuktikan bahwa CFIT-skala-2 memiliki daya diskriminasi yang baik dalam membedakan individu dengan kecerdasan tinggi dan rendah berdasarkan analisis IRT, sehingga dapat menjadi bukti relevansi administrasi CFIT-skala-2 yang masih bisa dilakukan hingga sekarang.

Namun demikian, pada penelitian ini CFI dan TLI masih tergolong rendah. Meskipun validitas konstruk menunjukkan model fit yang cukup baik, rendahnya nilai CFI dan TLI menunjukkan bahwa ada faktor yang belum sepenuhnya terakomodasi dalam model pengukuran.

Dibandingkan dengan penelitian Mustari (2015) yang menganalisis CFIT-skala-3, hasil penelitian ini menunjukkan kelemahan yang serupa dalam daya diskriminasi dan efektivitas item. Namun, penelitian lain seperti Hays & Smith (1980) menemukan bahwa CFIT memiliki korelasi yang cukup kuat dengan tes kecerdasan lainnya seperti WISC dan WAIS, sehingga dianggap mampu mengungkap konstruk psikologis yang diukur yaitu kecerdasan.

Relevansi penelitian

Dalam konteks asesmen psikologi di Indonesia, penggunaan alat ukur yang valid dan reliabel seperti CFIT-skala-2 sangat krusial, terutama mengingat keragaman budaya dan bahasa di Indonesia. Penelitian ini memberikan kontribusi penting dengan mengidentifikasi area yang memerlukan perbaikan pada alat ukur, seperti struktur item, daya diskriminasi, dan efektivitas distraktor. Temuan ini sejalan dengan studi terbaru oleh Fadillah (2021), yang menekankan perlunya evaluasi

mendalam terhadap properti psikometri CFIT untuk memastikan akurasi dan keadilan dalam pengukuran kecerdasan di Indonesia

Selain itu, penelitian oleh Suwandi (2015) juga menyoroti pentingnya uji invariansi pengukuran pada CFIT untuk memastikan bahwa tes ini bebas dari bias budaya dan dapat diterapkan secara adil di berbagai kelompok masyarakat. Oleh karena itu, hasil penelitian ini sangat relevan dan mendesak untuk dijadikan dasar dalam upaya revisi dan adaptasi CFIT-skala-2 agar lebih sesuai dengan konteks budaya dan kebutuhan asesmen di Indonesia.

Pendapat mahasiswa mengenai artikel yang dianalisis.

Sebagai mahasiswa psikologi, artikel ini memberikan wawasan yang mendalam tentang properti psikometri alat ukur psikologis, termasuk yang paling penting adalah validitas dan reliabilitas, khususnya dalam asesmen kecerdasan. Salah satu kekuatan artikel ini adalah pendekatan yang digunakan dalam menganalisis alat tes secara komprehensif dengan teori tes klasik dan teori respons butir. Namun, penelitian ini juga menunjukkan adanya kekurangan dalam model fit, yang seharusnya menjadi perhatian bagi pengembang alat tes di masa depan.

Selain itu, artikel ini menyoroti bahwa meskipun CFIT-skala-2 diklaim sebagai tes kecerdasan yang bebas budaya, masih terdapat keterbatasan dalam penerapannya di masa sekarang. CFIT-skala-2 menunjukkan variasi tingkat kesulitan yang tidak terstruktur secara sistematis. Daya beda butir tidak optimal, dan beberapa distraktor tidak efektif. Meskipun CFIT-skala-2 tidak mengandung DIF, struktur penyusunan butirnya masih perlu diperbaiki. Analisis CFA menunjukkan bahwa CFIT-skala-2 mengukur konstruksi teoretis yang sesuai. Namun, reliabilitas konstruk masih perlu ditingkatkan. Penelitian ini memberikan pemahaman yang lebih dalam tentang bagaimana suatu alat ukur perlu dievaluasi dan disesuaikan sebelum digunakan dalam berbagai konteks sosial dan budaya. Pengembangan lebih lanjut dapat dikaji, utamanya di Indonesia.

Implikasi Hasil Jurnal terhadap Bidang Asesmen Psikologi

Hasil penelitian ini memiliki implikasi yang luas terhadap bidang asesmen psikologi, khususnya dalam pengembangan dan revisi alat ukur kecerdasan. Salah satu implikasi utama adalah perlunya revisi terhadap CFIT-skala-2 agar lebih sesuai dengan populasi yang diuji. Dalam praktik asesmen psikologi, validitas dan reliabilitas merupakan faktor utama yang menentukan sejauh mana suatu tes dapat memberikan hasil yang akurat dan konsisten. Jika alat tes seperti CFIT-skala-2 memiliki kelemahan dalam daya diskriminasi dan efektivitas distraktor, maka hasil asesmen yang diperoleh bisa menjadi kurang akurat dan dapat mempengaruhi pengambilan keputusan di berbagai bidang, seperti pendidikan dan seleksi kerja.

Selain itu, penelitian ini juga menekankan pentingnya penggunaan pendekatan *multimethod* dalam asesmen psikologi. Dengan menggabungkan teori tes klasik dan teori respons butir, asesmen psikologi dapat menjadi lebih akurat dan objektif. Oleh karena itu, hasil penelitian ini dapat menjadi dasar bagi pengembangan alat tes kecerdasan yang lebih adaptif dan relevan dengan kebutuhan masyarakat Indonesia. Penelitian ini menyarankan agar karakteristik psikometrik CFIT-skala-2 direvisi dan dievaluasi ulang untuk meningkatkan kualitas pengukuran kecerdasan individu.

Daftar Pustaka

Fadillah, F. Analisis Peluang Guessing pada Culture Fair Intelligence Test (Cfit) 3A dengan Metode IRT.

Suwandi, B. (2015). *Uji measurement invariance pada culture fair intelligence test menggunakan pendekatan multiple-group confirmatory factor analysis* (Bachelor's thesis, Fakultas Psikologi UIN Syarif Hidayatullah Jakarta).

Zahra, N. S., Widyastuti, & Ridfah, A. (2022). *Psychometric Characteristics of the Culture Fair Intelligence Test Scale 2*. JP3I (Jurnal Pengukuran Psikologi dan Pendidikan Indonesia), 11(2), 123-136. DOI: http://dx.doi.org/10.15408/jp3i.v10i1.25947